

Collegio Carlo Alberto



Import Diversification along the Growth Path

Esteban Jaimovich

No. 242

February 2012

Carlo Alberto Notebooks

www.carloalberto.org/working_papers

© 2012 by Esteban Jaimovich. Any opinions expressed here are those of the authors and not those of the Collegio Carlo Alberto.

Import Diversification along the Growth Path^{*}

Esteban Jaimovich

Collegio Carlo Alberto

February 2012

Abstract

We provide evidence showing that the degree of diversification of import sources of finely disaggregated commodities rises monotonically along the growth path. This result is robust to different measures of import diversification and the inclusion of a large set of additional control variables. In addition, we show the process of rising import diversification takes place as countries gradually increase their spending shares in imports originating from relatively distant exporters.

Key Words: Bilateral trade, import diversification, product variety.

JEL Codes: F10, F14.

^{*}I would like to thank Ainhoa Aparicio-Fenoll and Alina Sandor for useful comments.

1 Introduction

The recent literature on bilateral trade flows has established a peculiar empirical regularity: countries tend to purchase their imports from relatively few available sources.¹ As a striking example, Haveman and Hummels (2004) find that, when defining commodities by the 4-digit Standard International Trade Classification (SITC) categorisation, 58% of importers buy their imported commodities from less than 10% of the available exporters. In other words, bilateral trade flows data is plagued with zeros.

The number of zeros, although ubiquitous and large in relative terms, has not remained constant across time. Broda and Weinstein (2006) and Sauré (2011) found that the ‘zeros’ in the bilateral trade data display a downwards trend over time – the former article looks only at US data while the latter looks at cross-country data. Such increase in the number of trading partners has been deemed a side-effect of the secular rise in per capita incomes in the world. The argument is the following: as real incomes grow, consumers’ love-of-variety rises. Wealthier individuals then widen the set of varieties they consume, which in turn leads to an expansion of the sources/exporters from where they purchase their goods.

These stylised facts concerning the ‘zeros’ in the bilateral trade data are certainly interesting in their own right. However, they fall short of giving a thorough characterisation of the degree of diversification of imports, and how import diversification varies with the level of income of the importer. In particular, the discreteness implied by treating trade links as a binary variable (that is, either we observe ‘positive trade’ or ‘zero trade’) disregards large amounts of useful

¹E.g., Haveman and Hummels (2004), Besedeš and Prusa (2006), Broda and Weinstein (2006), Helpman, Melitz and Rubinstein (2008), Baldwin and Harrigan (2011), Chen (2012).

information contained in the data.

In this paper, we contribute to the above-mentioned literature by studying the evolution of import diversification along the growth path. We use two different measures of import diversification at the product level: (i) the Gini coefficient for value of imports by source; (ii) the (normalised) Herfindahl index for import shares by source. Both measures of diversification, which are described in further detail in the next section, make use of all the available information in terms of total value of imports by source.²

In terms of results, Section 3 shows that import diversification increases monotonically along the growth path. This results remains robust to the inclusion of a large set of additional controls, such as product, importer, product-importer, and time dummies. In addition, in Section 4, we provide some suggestive evidence concerning the geographical pattern of import diversification: the data suggests that imports diversification takes place as countries shift part of their (import) spending shares from nearby exporters to more distant ones.

2 Data and Measures of Import Diversification

Our main data source is Feenstra et al (2005). This dataset compiles bilateral trade data by commodities for the years 1962-2000. Traded commodities are organised according to the 4-digit SITC (revision 2). In total, there are 1004 different commodities in the SITC-4 classification. We use the data on value of imports of each SITC-4 commodity (by country of origin

²Imbs and Wacziarg (2003) have previously used the Gini coefficient and the Herfindahl index to measure the degree of concentration of labour and value added across different sectors in the economy. In this paper, we apply a similar methodology, but we use it to measure the degree of diversification of imports of finely disaggregated goods in terms of country of origin of those imports.

of the commodity) to compute two different measures of import diversification by commodity-importer-year: the Gini coefficient for inequality of sources of imports, and the normalised Herfindahl index for concentration of import shares. Both measures are constructed using the value of imports of each SITC-4 commodity in a given year.

To compute the Gini coefficients, we first order, from smallest to largest, the total value of imports of each commodity by export source, considering all active exporters of the commodity in a given year. Notice this includes the cases when purchases by an importer from an exporter in a given year are actually equal to zero. Next, after ordering the value of imports by source, we apply the following formula:

$$GINI_{i,m,t} = 2 \frac{\sum_{x=1}^{N_{i,t}} x V_{m,i,t,x}}{N_{i,t} \sum_{x=1}^{N_{i,t}} V_{m,i,t,x}} - \frac{1 + N_{i,t}}{N_{i,t}}. \quad (1)$$

$GINI_{i,m,t}$ yields the value of the Gini coefficient of imports of commodity i , by importer m , in year t . $V_{m,i,t,x}$ denotes the value of imports of product i , by importer m , in year t , originating from exporter x – each exporter x has been *numerically* ordered within the sum in (1) according to their value of exports of i to m in t , from smallest to highest, including those who exported zero. $N_{i,t}$ is the total number of *active* exporters of commodity i in period t : an active exporter of i in t is a country that has exported a strictly positive amount of good i at least to one importer in year t .

The computation of the normalised Herfindahl indices is more straightforward. Denoting by $HERF_{i,m,t}$ the value of the normalised Herfindahl index of import shares of commodity i , by importer m , in year t , we have:

$$HERF_{i,m,t} = \frac{\sum_{x=1}^{N_{i,t}} \left(\frac{V_{m,i,t,x}}{\sum_{x=1}^{N_{i,t}} V_{m,i,t,x}} \right)^2 - N_{i,t}^{-1}}{1 - N_{i,t}^{-1}}. \quad (2)$$

Notice that each term $V_{m,i,t,x} / \sum_{x=1}^{N_{i,t}} V_{m,i,t,x}$ in (2) equals the market share of exporter x in country-importer m , for the case of commodity i during period t .³

Lower values of either $GINI_{i,m,t}$ or $HERF_{i,m,t}$ indicate a more diversified provision of good i in period t in terms of sources, given the total number of active exporters of i in t ($N_{i,t}$).

The rest of the data we use include the following variables: income per capita (in purchasing power parity) and population taken from Heston, Summers and Aten (2011), and the distance (in kilometers) between exporter and importer obtained from Mayer and Zignano (2006).

Table 1 reports some summary statistics for the income per capita, Gini, normalised Herfindahl and number of active exporters ($N_{i,t}$). In the dataset there are 5,940 observations with a Gini coefficient equal to 0. These extreme observations essentially correspond to the cases in which $N_{i,t}$ equals 1 (i.e., there exists only one active exporter of good i in t).⁴ In these peculiar cases the normalised Herfindahl index is not well-defined (so, we lose these observations in the regressions involving $HERF_{i,m,t}$). We should stress that our regressions in Section 3 will control for the number of active exporters by including one dummy variable for each of the possible values taken by $N_{i,t}$.⁵ As a final remark, it might look striking the fact that the Gini coefficients lie so overly concentrated near unity. This is the consequence of the relatively large presence of zero trade between active exporters and importers in the dataset.

³We have also run all the regression displayed in Section 3 for the case of the (standard) Herfindahl index: $\sum_{x=1}^{N_{i,t}} \left(V_{m,i,t,x} / \sum_{x=1}^{N_{i,t}} V_{m,i,t,x} \right)^2$. The regression results using the Herfindahl (which are available from the author upon request) are very similar to those using (2).

⁴More precisely, there are 5936 observations (representing 0.24% of all observations) in which $N_{i,t} = 1$; the other four cases in which $GINI_{i,m,t} = 0$ correspond to observations with $N_{i,t} = 2$.

⁵An alternative to this, though more demanding in terms of total number of regressors, would be to include time-product dummies.

Table 1: Summary Statistics

	mean	median	st. dev.	min	max
GDP pc (in 1,000 PPP US dollars)	7.57	3.76	9.00	0.118	74.3
Gini	0.953	0.966	0.059	0	0.993
Normalised Herfindahl	0.547	0.488	0.309	0	1
Number of active exporters ($N_{i,t}$)	54.94	54	29.9	1	160

Number of observations: Gini 2,468,374; normalised Herfindahl 2,462,438

3 Import Diversification and Growth

This section reports our main empirical results. Tables 2.A and 2.B present a series of regressions with $GINI_{i,m,t}$ and $HERF_{i,m,t}$ acting as dependent variable, respectively. Except for column (4), all regressions contain the *logarithm* of income per capita as the independent variable of main interest; in column (4) we use instead the *level* of income per capita. In all regressions we add the level of population as independent variable, in case the overall size of the economy may somehow influence the extent of diversification of imports (for example, due to the presence of sunk costs in the creation of trade links). In addition, we always include time dummies; these would control for any possible common trends in import diversification (for example, owing to time trends in transport costs), or the common effect of some major global shock impacting the intensity of trade (for example, an oil shock).

In column (1) of Table 2.A, we regress the Gini coefficient on the aforementioned variables, including also product and importer dummies. The former would take care of the facts that different commodities are generally offered by different sets of exporters.⁶ The latter would control for time-invariant importer characteristics, such as area, climate diversity, being landlocked or

⁶In particular, the product dummies would control for the bias that may be generated if richer countries tend to allocate a larger share of their spending on goods offered in general by a larger set of exporters, for which there is thus greater scope for diversification of export sources.

an island, etc. The regression shows that the log income per capita correlates negatively with the value of the Gini coefficients. This result suggests that, as economies become richer, they tend to purchase their imports of finely disaggregated products from a more diversified set of exporters of those products.

In column (2) we introduce a dummy variable for each of the possible values taken by $N_{i,t}$ in the dataset. We do so in case the (time-invariant) product dummies alone turn out to be not enough to fully control for the possible bias generated by the presence of heterogeneous set of exporters across products and time (this may actually occur if there is substantial variation of the value of $N_{i,t}$, *within* products, *across* time). As we can observe, the results in column (2) are essentially identical to those in column (1), although the R squared does rise substantially.

The regression in column (3) includes a full set of importer-product dummies. Importer-product dummies are able to control for some additional features not taken care of by (1) or (2), such as the relative distances of a specific importer from the main export sources of a specific product, or the level of self-sufficiency of the country in a specific product. Although slightly smaller (in absolute value) than in columns (1) and (2), our main correlation of interest remains negative and highly significant. As a robustness check, column (4) reports the results of the same regression, but using the level of income per head instead of its logarithm – the negative and significant correlation remains in place.

As additional robustness check, in columns (5) and (6) we split the sample and run two separate regressions with the same set of independent variables used in (3). In the former we use only the observations for which $N_{i,t} < 56$, while in the latter we include only those for which $N_{i,t} > 56$. That is, we take the sub-samples in which the number of active exporters is below and above the sample median, respectively. The rationale for this is that our correlation

of interest may vary significantly at different levels of $N_{i,t}$, since the scope for increasing the level import diversification in a given product will be somehow conditioned by $N_{i,t}$. The results in (5) and (6) yield two main insights. First, in both sub-samples the log of income per head and the Gini are negatively correlated. Second, although the partial correlation coefficient in (6) is slightly larger than in (5), their difference is statistically not significant.

All the previous regressions are conducted under the assumption that the correlation between import diversification and income per head remains the same at all levels of per capita income. However, one may worry about the validity of this assumption. In particular, one may want to rule out the possibility of a non-monotonic relationship between the two variables. In columns (7) - (10) we proceed to approach this issue. In order to do so, we split the dataset into four sub-samples determined by the income per capita quantiles, and run separate regressions for each of the quantiles. The correlation of interest is negative across the board. We do, however, find an interesting pattern concerning the estimates: the association between the log of income per head and import diversification becomes steeper as economies get richer; in other words, the process of import diversification speeds up along the growth path.⁷

Finally, Table 2.B reports the results obtained when using the normalised Herfindahl as the dependent variable in the regressions. As we can observe, all the main results remain essentially intact under this alternative measure of import diversification.⁸

⁷The per capita income quantiles are: $Q1 = [0; 1180]$, $Q2 = [1180; 3760]$, $Q3 = [3760; 10170]$, and $Q4 = [10170; 74300]$. Similar results (available from the author upon request) are obtained when we split the sample in ten sub-samples according to income deciles.

⁸The only qualitative difference appears in the quantile regressions – columns (7) - (10). When diversification is measured using the normalised Herfindahl index, the speed of diversification seems to slightly slow down along the growth path (although the differences in the point estimates are far from being significant).

Table 2.A - Dependent Variable: Gini coefficient by good and importer (multiplied by 100)

	Full Sample				$N_{i,t} < 56$		$N_{i,t} > 56$		Income per capita split by quantiles			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
Log income per capita	-1.430 (0.126)***	-1.454 (0.130)***	-1.304 (0.137)***		-1.251 (0.177)***	-1.363 (0.126)***	-0.716 (0.104)***	-0.912 (0.187)***	-1.054 (0.378)***	-1.622 (0.481)***		
Population	-0.053 (0.011)***	-0.056 (0.011)***	-0.053 (0.008)***	-0.091 (0.009)***	-0.056 (0.007)***	-0.050 (0.012)***	-0.013 (0.006)***	-0.109 (0.032)***	-0.270 (0.109)***	-0.040 (0.064)		
Income per capita				-0.100 (0.017)***								
Year dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES		
Product dummies	YES	YES	-	-	-	-	-	-	-	-		
Importer dummies	YES	YES	-	-	-	-	-	-	-	-		
Importer-Product dummies	-	-	YES	YES	YES	YES	YES	YES	YES	YES		
Dummies for $N_{i,t}$	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES		
Observations	2,334,149	2,334,149	2,334,149	2,334,149	816,834	1,517,315	419,801	536,073	638,534	739,785		
R squared	0.36	0.85	0.92	0.91	0.95	0.77	0.95	0.94	0.95	0.85		

Table 2.B - Dependent Variable: normalised Herfindahl index by good and importer (multiplied by 100)

	Full Sample				$N_{i,t} < 56$		$N_{i,t} > 56$		Income per capita split by quantiles			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)		
Log income per capita	-13.478 (1.33)***	-13.767 (1.32)***	-13.300 (1.45)***		-12.24 (1.59)***	-13.98 (1.45)***	-12.829 (1.85)***	-12.240 (2.51)***	-12.416 (3.71)***	-10.429 (3.78)***		
Population	-0.504 (0.091)***	-0.502 (0.089)***	-0.499 (0.091)***	-0.836 (0.171)***	-0.485 (0.091)***	-0.471 (0.101)***	-0.170 (0.102)*	-0.298 (0.288)	-2.245 (0.966)**	-0.862 (0.580)		
Income per capita				-0.687 (0.135)***								
Year dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES		
Product dummies	YES	YES	-	-	-	-	-	-	-	-		
Importer dummies	YES	YES	-	-	-	-	-	-	-	-		
Importer-Product dummies	-	-	YES	YES	YES	YES	YES	YES	YES	YES		
Dummies for $N_{i,t}$	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES		
Observations	2,328,503	2,328,503	2,328,503	2,328,503	811,188	1,517,315	418,854	534,732	635,960	738,963		
R squared	0.41	0.43	0.63	0.63	0.66	0.64	0.53	0.60	0.62	0.70		

Robust absolute standard errors clustered at the importer and product level in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%. Data is for years 1962-2000. Number of SITC-4 products: 1004. Number of importers: 160. Income per capita is measured in thousands of 2005 US dollars (PPP adjusted). Population is measured in 10 million people.

4 Diversification and Distant Markets

Section 3 remains silent as to whether there is some general pattern regarding how the diversity of import sources by commodity rises along the growth path. In this section we present some results suggesting that import diversification goes hand in hand with a rise in import shares originating from relatively distant markets. In a sense, such relationship between diversification and distance sounds theoretically intuitive: it suggests that love-of-variety leads individuals to expand the set of varieties they consume by gradually adding commodities originating from less accessible markets.

We compute the average distance of imports of commodity i ($AvgDist_{i,m,t}$) by averaging the bilateral distance between exporter x and importer m ($Dist_{m,x}$), where the variable $Dist_{m,x}$ is weighted by the share of imports of i originating from x . Namely,

$$AvgDist_{i,m,t} = \sum_{x=1}^{N_{i,t}} \left(Dist_{m,x} \times \frac{V_{m,i,t,x}}{\sum_{x=1}^{N_{i,t}} V_{m,i,t,x}} \right).$$

In Table 3 we conduct a series of regressions with $AvgDist_{i,m,t}$ acting as dependent variable. All regressions include year dummies, importer-product fixed effects, a set of dummies for each of the possible values taken by $N_{i,t}$ in the dataset, and the importer's population.

The regressions in columns (1) and (2) contain only one additional regressor each: in the former this is the logarithm of income per head, while in the latter the Gini coefficient. The regression in (1) yields a positive coefficient for the income variable, although the level of significance does not reach 10% (the result would be significant at 10% if the test were conducted only against the upper tail). On the other side, the regression in (2) produces a positive coefficient on the Gini, and highly significant.

Next, in column (3), we run a regression including simultaneously the log income per head

and Gini. Interestingly, the Gini keeps its point estimate and level of significance essentially intact. By contrast, the coefficient associated to income per head falls quite drastically, and as a result it turns highly non-significant. From columns (4) and (5) we may observe that the same qualitative results also arise when using the normalised Herfindahl to measure import diversification (moreover, the drop in the point estimate of the coefficient associated to log income per capita is quantitatively more pronounced when adding $HERF_{i,m,t}$ to regression (1) than when adding $GINI_{i,m,t}$ to it).

Taken all together, the regressions in Table 3 point towards a particular geographical pattern of diversification of import sources. The rise in import diversification seems to take place as more distant exporters increase their market shares along the importer's growth path. In addition, the regressions appear to rule out the presence of a strong "direct" income effect, whereby a rise income per head allows the substitution of nearby exporters with more distant ones. Instead, it seems to be the case that a higher income per capita leads to a rise in the average distance of import sources through its (positive) effect on import diversification. This result is consistent with the idea that love-of-variety leads individuals to expand the set of varieties they consume by gradually adding commodities originating from less accessible markets.

5 Conclusion

Identifying commodities according to the 4-digit SITC classification, we have shown that the degree of import diversification by products in terms of suppliers rises monotonically along the growth path. This result is robust to measuring diversification either by means of the Gini coefficient or the Herfindahl index. It is also robust to the inclusion of a large set of additional

Table 3 - Dependent Variable: Average distance of imports (weighted average)

	(1)	(2)	(3)	(4)	(5)
Log income per capita	199.4 (140.7)		135.2 (139.5)		110.3 (136.1)
Gini \times 100		-47.4 (12.47)***	-49.2 (12.53)***		
Normalised Herfindahl \times 100				-6.54 (1.13)***	-6.66 (1.14)***
Population (in 10,000,000)	-27.7 (5.87)***	-27.7 (5.54)***	-30.4 (5.79)***	-28.5 (5.35)***	-30.6 (5.72)***
Year dummies	YES	YES	YES	YES	YES
Importer-Product dummies	YES	YES	YES	YES	YES
Dummies for $N_{i,t}$	YES	YES	YES	YES	YES
Observations	2,334,119	2,468,282	2,334,149	2,462,346	2,328,503
R squared	0.71	0.71	0.71	0.71	0.72

Robust standard errors clustered at the importer and product level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.
Number of products: 1004. Number of importers: 160. Data is for years 1962-2000. The dependent variable is constructed using distances from exporter and importer obtained from Mayer and Zignano (2006) weighted by the import share originating from each of the exporters.

controls. In addition, we provided some evidence suggesting that import diversification materialises as countries gradually shift part of their spending shares from relatively nearby exporters to more distant ones, consistent with the view of individuals' preferences displaying love of variety.

References

- [1] Baldwin, R. and Harrigan, J. (2011). 'Zeros, quality and space: trade theory and trade evidence', *American Economic Journal: Microeconomics*, vol. 3, pp. 60-88.
- [2] Besedeš, T. and Prusa, T. (2006). 'Ins, outs, and the duration of trade', *Canadian Journal of Economics*, vol. 39, pp. 266-295.
- [3] Broda, C. and Weinstein, D. (2006). 'Globalization and the gains from variety', *Quarterly Journal of Economics*, vol. 121, pp. 541-585.
- [4] Chen, Wei-Chih (2012). 'Innovation and duration of exports', *Economics Letters*, vol. 115, pp. 305-308.
- [5] Feenstra, R., Lipsey, R., Deng, H., Ma, A. and Mo, H. (2005) 'World trade flows: 1962-2000', NBER Working Paper No. W11040.
- [6] Haveman, J. and Hummels, D. (2004). 'Alternative hypotheses and the volume of trade: the gravity equation and the extent of specialization', *Canadian Journal of Economics*, vol. 37, pp. 199-218.
- [7] Helpman, E., Melitz, M. and Rubinstein, Y. (2008). 'Estimating trade flows: trading partners and trading volumes', *Quarterly Journal of Economics*, vol. 123, pp. 441-487.
- [8] Heston, A., Summers, R. and Aten, B. (2011). *Penn World Table Version 7.0*. Center for International Comparisons of Production, Income and Prices (University of Pennsylvania).
- [9] Imbs, J. and Wacziarg, R. (2003). 'Stages of diversification', *American Economic Review*, vol. 93, pp. 63-86.

- [10] Mayer, T. and Zignago, S. (2006). 'GeoDist: the CEPII's distances and geographical database', MPRA Paper 31243.
- [11] Sauré, P. (2011). 'Bounded love of variety and patterns of trade', *Open Economies Review*, forthcoming.